Journal of Applied Computing Research, 1(1):20-32
January-June 2011

Recommending knowledge in a knowledge based social network[1]

Edeilson M. Silva, Ricardo A. Costa, Paulyne Jucá
Universidade Federal de Pernambuco. Caixa Postal 7851, Cidade Universitária.
50670-901, Recife, PE, Brasil. ems5@cin.ufpe.br, rac@cin.ufpe.br, pmj@cin.ufpe.br

Lucas Roberto Bechert Schmitz, Silvio R.L. Meira
Centro de Estudos e Sistemas Avançados do Recife - C.E.S.A.R. Rua Bione, 220, Bairro do Recife
50030-390, Recife, PE, Brasil. lucas.schmitz@cesar.org.br, srlm@cin.ufpe.br

Abstract: The organizations aim to increase its competitiveness. In this context, they have been searching for new ways to improve their productivity, the quality of their products, and cost reduction. To achieve these goals, it is essential to use the collaborators' potentials and the relationship among them to find and share tacit knowledge. Since tacit knowledge is stored in people's mind, it is hard to be formalized and documented. Facing this difficulty, identifying and recommending persons who retain the needed knowledge might be a good option. This work presents the Specialist Recommender System (SWEETS) and its application into the a.m.i.g.o.s. environment, a social network platform for knowledge management. The SWEETS system uses folksonomy to extract a lightweight ontology, which is essential to effectively identify people's skills. This lightweight ontology is based by tags (concepts) relating them to items (instances), and its co-occurrences. In addition, such ontology is domain independent, which is a contribution of this work. Applying the SWEETS system into the a.m.i.g.o.s. environment we are looking for minimizing the communication problem in the corporation, providing an improvement on knowledge sharing. Therefore, a better usage of the collaborators knowledge may be expected.

Key words: SWEETS, social network, knowledge management.

Introduction

According to a study from the IBM[2], CEOs around the world are interested in establishing an environment and culture to provide support to help their companies to innovate. For this, the CEOs need to exploit the maximum potential of their companies. This potential is directly related to the knowledge of its employees and the relationship between them. That is, the ability to find and share tacit knowledge. In this scenario, providing quick access to the knowledge classification is essential for companies that aim to avoid duplicate efforts and innovate.

According to estimates released by Duhon (1998), from 50% to 90% of the corporate knowledge is in the minds of their employees – tacit knowledge, which is hard to formalize and document (Nonaka and Takeuchi, 1995). Because of this difficulty, it would be a really interesting approach to provide an environment that people can use to naturally formalize their knowledge. Thus, this formalized knowledge could be essential for companies that aim to avoid duplicate efforts.

Furthermore, a common and well known practice is the search of information among people from the same group. That is because peo-

[1] Article based on masters dissertation of Edeilson Milhomem Silva, *SWEETS: um sistema de recomendação de especialistas aplicado a redes sociais* (2009). A similar article was published in Portuguese: "SWEETS: um Sistema de Recomendação de Especialistas aplicado a uma plataforma de Gestão de Conhecimento" in *Revista de Informática Teórica e Aplicada*, Vol. 18, N. 1, p. 83-111. Available at: http://seer.ufrgs.br/rita/article/view/rita_v18_n1_p83/11824.
[2] IBM Global Business Services, Global CEO Study 2006. Available from: http://www.ibm.com/bcs/ceostudy.

ple tend to give more credibility to information from their own social networks - colleagues and friends (Nardi *et al.*, 2000; Bogartti and Cross, 2003; Plickert *et al.*, 2007). According to the analysis realized at Casonato and Harris (2002), the employees of an organization acquire 50 to 75% of their information directly from other people.

Although personal networks are a way to get quick answers, sometimes they are not sufficient to get the people who have knowledge about a subject. Thus, those social networks have a limited range. People in a personal network can be intermediaries, facilitating the contact with people that do not know each other (Ehrlich *et al.*, 2007), and so, provide greater interactivity, communication and collaboration.

Within a social context, if someone has a question to ask, it is more likely that this question will be answered by a friend instead of an unknown-person. Thus, according to Ehrlich *et al.* (2007), a system that identifies such experts could use the social context in which people is embedded to make such recommendations. Still, according to Ehrlich *et al.* (2007), researchers argue that any social network-based system should provide technologies to search experts. Thus, finding people who are capable of solving a problem would be helpful to users of this kind of system.

This paper presents an expert recommender system and its implementation on a knowledge management platform a.m.i.g.o.s. (Environment for Integration of Social Groups and Organizations), currently used on C.E.S.A.R. - Center for Advanced Studies and Systems of Recife - Brazil.

This work is organized as follows: first it presents concepts related to Social Networks, then it approaches some characteristics of the Knowledge Management area, after that it presents some expert recommender systems already available in the literature, then it shows the proposed expert recommender system, the next section presents the case study, then it presents the experiments and results, and finally it presents the final considerations.

Social networks

The Social Networks theory approaches the social relationships as nodes and links. Each node represents an actor within the social network, and each link represents a social connection between actors. There are many different ways to link these actors, each one related to the nature of the represented social network (Iacobucci *et al.*, 1994).

Works in social networks are as old as human history, but, only in the last decades, people have noticed social networks as an organizational tool. According to Lipnack and Stamps (1988), what is new in working with network of connections is the promise of a global organization focusing on individual participation. The work in network of connections can lead to a global perspective based on individual experience.

Since the mid 90's, the Web Based Social Networks (WBSN) evolved and proliferated in a fast pace, both in number of WSBN and scope. There is a really large set of WSBN, each with its defined scope, ranging from business and entertainment to pet relationships.

Great part of the success achieved by WBSN is due to the convenience provided by some of them to promote the interaction and communication between its members. Such conveniences can also be provided by virtual communities that, despite having some similarities, are not exactly a synonym to social networks.

The term virtual community was first coined in 1987, referring to the emergence of social gatherings through computer-mediated communication (Boomen, 2008). There are many different variations of the definition of a virtual community; some authors define a virtual community as any group of people interacting with each other via communication channels such as the Internet (Gal-oz *et al.*, 2008), while others define it as a social aggregation on the Internet, having a core of recurrent users, engaged in ongoing group interactions at a shared virtual place, whether norms and values could arise from the aggregation, invoking a sense of belonging (Gal-oz *et al.*, 2008).

According to these definitions, it is not correct to affirm that a Web Based Social Network is a synonym to a Web Based Community. It is possible to conclude that the essence of a virtual community is on communication and sense of belonging, which are not necessarily present in every WBSN. In particular, members of social networks like Facebook or MySpace, where the main attraction is entertainment, are not communicating to other users, or do not feel a sense of belonging to the so called social network community.

But some WBSN do have virtual community characteristics, as the LinkedIn and Orkut social networks. In these networks there are specific features for community creation inside the social network, and the communication tools allow recurrent users discuss about specific topics. In this case we can affirm that there are virtual communities inside the WBSN.

But there are other examples of WBSN which can be classified as virtual communities. Lately, the Corporative Social Networks, which consist of private Social Networks accessible only to company employees,have been used as communication and collaboration tools for members of an organization. In these cases it is possible to affirm that there is a core of recurrent users which are engaged in group interactions and also have a sense of belonging to the company social network.

One of the reasons justifying the interest of organizations in social networks, specially the ones which can also be called as virtual communities, is how efficient these networks are to share the knowledge inherent to each person (Staab *et al.*, 2005), that is, the tacit knowledge defined by Polanyi (1967), which in an organization context is also known as know-how (Nonaka and Takeuchi, 1995). Once that knowledge, that is relevant to the members of the Social Network, is documented, it can be reused, avoiding wasting of effort from employees. The possibility of summarizing the knowledge is directly related to better exploiting the existing human capital in a given organization.

According to Domingos and Richardson (2001), the way that users publish information in social networks is impressive and without any precedent. Therefore, these networks can also be seen as a huge data repository, which contains relevant information about every connected user. Such information may be used for a variety of purposes, such as an application that infers the confidence or degree of trust of each user.

Besides that, the usage of social networks also permits to find some unusual correlations, which, according to Staab *et al.* (2005), may allow the discovering of new information about quantitative and qualitative aspects of the social connections. Thus, this kind of environment can be viewed as an excellent way to accomplish analysis and research related to knowledge flow inside an organization.

Because the use of social networks is an efficient way to share and distribute individual knowledge, it becomes an interesting approach to support a knowledge management initiative. In 2000, Erickson and Kellogg (2000) began working towards the development of a multi-use environment, which allowed communication and collaboration in groups, where communitarian knowledge might be created.

The knowledge management methods can focus on both tacit and explicit knowledge. Making an analysis of circumstances in which this knowledge can fit in a social network, it was found that:

- Explicit knowledge is directly related to creation, storage, sharing and use of explicitly documented knowledge (Hansen *et al.*, 1999) and therefore is directly related to social networks in many aspects, such information posted in the communities, information interchange supported by chat tools and files associated to communities.
- Tacit knowledge is directly related to inter-personal relationships (Hansen *et al.*, 1999), in other words, relationships that a user has within the Social Network. Thus, all community knowledge, resulting of the interactions between users, can be characterized as tacit knowledge.

Moreover, other benefits may be related to the use of social networks, such as being an interactive and generally informal environment, so users can express their thoughts in a natural and more freely way, enriching the organization memory. The success of WBSN is attracting a lot of attention, turning into an increased level of interest in this area, and consequently providing new directions to the social networks research area.

Knowledge management

According to Choi and Lee (2003), Knowledge Management (KM) in a software corporation is an opportunity to create a default perception language among software developers so that they can interact, deal and share knowledge and experiences. The reduction in loss of Intellectual Capital from employees who leave the company; the cost reduction for the development of new products; and the increased productivity by making knowledge easily accessible to all employees are some of several benefits of the use of a Knowledge Management strategy. The KM area is receiving special attention from companies searching for competitive advantage. The following statistics are a simple proof of this attention (Bose, 2004).

- A total of 80 percent of Fortune 500 companies have KM staff;
- Texas Instruments has saved $1 billion since it launched KM programs in the mid-1990s;
- Of CEOs polled at the 2001 World Economic Forum in Davos, Switzerland, 95 percent said that KM was critical to organizational success;
- Of Canadian business leaders polled by IpsosReid in 2001, 91 percent believed that KM practices have a direct impact on organizational effectiveness.

Polanyi (1967) categorized knowledge in tacit and explicit. The latter is basically what can be easily documented and distributed, while the tacit knowledge resides in the human mind, behavior and perception, and thus is difficult to be formalized and distributed. As expected, the traditional knowledge managements approaches usually focuses on explicit knowledge, but some authors indicate that it is necessary to capture, process and transfer tacit knowledge in order to fully understand an organization process. Identify people who have knowledge about a subject is a way to find tacit knowledge (individual knowledge).

Expert recommender systems

An Expert Recommender System is an alternative to provide access to the implicit/tacit knowledge – that is in people's mind. In this sense, the literature presents some initiatives to provide access to the tacit knowledge. Among these initiatives, we can mention ReferralWeb (Kautz *et al.*, 1997), ERS (Yukawa *et al.*, 2002), TABUMA (Reichling *et al.*, 2005), ICARE (Petry, 2007) and SmallBlue (Lin *et al.*, 2008).

ReferralWeb (Kautz *et al.*, 1997) is an expert recommender system that combines concepts inherent to Social Networks and Collaborative Filtering to provide personalized recommendations, prioritizing those specialists closest to the user, that is, those whose social distance is smaller. The relationships among people were extracted from email logs, because the assumption that email would be a rich source of social relationships extraction (social network). However, this feature can be considered a problem because it raises important concerns related to information privacy.

ERS (Expert Recommender System) (Yukawa *et al.*, 2002) uses information retrieval methods to return people and/or organizations with strong relevance to a keyword or document. It uses a document base to find experts with relevance taking into account the desired topic and the person. The analysis of documents related to users which is made to infer whether a user is an expert in a given area can be a problem, because the efficiency of recommendations would be directly related the number and quality of these documents.

Just as ERS (Yukawa *et al.*, 2002), TABUMA (Reichling *et al.*, 2005) is an expert recommender system that exploits the ability of documents reflecting interests of a user. A set of documents associated to the user's work is used to generate his profile. This tool has the advantage of being very flexible because it can accept any type of document as input, which indicates skills and experiences of users. However, this freedom may become a disadvantage if the user does not submit documents reflecting his expertise.

ICARE (Petry, 2007) is a context-sensitive expert recommender system. It uses a domain's ontology to achieve recommendations. ICARE, recommends experts considering information as much of the user who makes the request, as of the experts that are recommended, based on keywords informed by users. In this way, it promotes personalized recommendations, due to the fact that they change according to the user and the time instant in which the request is made. That is, the experts more appropriate to offer assistance in a particular instant are recommended. Among other information, the availability of the specialist, the role that the expert occupies in the organization, the social distance among the target user and the expert and expert's reputation in the set of people that interact with him are used to determine which users should be recommended.

SmallBlue (Lin *et al.*, 2008) aims to find experts, communities and social networks in large companies, through data mining techniques, retrieval information and social networks analysis. The social network on SmallBlue is extracted from e-mail messages (similar to ReferralWeb) and instant messaging, that is, the user manages his information. This expert system also shares the same privacy problems of the ReferralWeb, as previously mentioned.

SWEETS

SWEETS 2.0 is a passive expert recommender system. It can be deployed to any computing environment having information associated to users and using folksonomy concepts. Folksonomy, according to Mika (2007), is a renewal linguistic for collaborative categorization, from the free use of keywords. That is, a mechanism of social tagging in which people collaborates to its creation. This collaboration is possible from descriptions of shared objects realized by users.

For Mika (2007), folksonomy appears as a good alternative to the emergence of lightweight ontology and creation of metadata. One of the advantages of using folksonomy for the extraction of lightweight ontologies is that a folksonomy can cover several areas, unlike a domain's ontology that represents knowledge about only one area. Thus, the use of lightweight ontologies emerged from a folksonomy, can be applied in many contexts.

Also according to Mika (2007), to make models of folksonomy networks in an abstract level, a system tripartite graphs can be used, where the set of vertices is defined in 3 (three) disjoint sets: $A = \{a_k, \ldots, a_{k-1}, a_k\}$, $C = \{c_1, \ldots, c_{l-1}, c_l\}$, $I = \{i_m, \ldots, i_{m-1}, i_m\}$. Each of these sets corresponds respectively to the actors (users), concepts (tags) and the instances noted (e.g., documents, websites, images). That way, a social tagging system allows users associate tag to objects, creating ternary associations among user, object and concept. Thus, a folksonomy (T) is defined by a set of notes $T \subseteq AxCxI$. From this, Mika (2007) affirms to be possible to use the traditional bipartite model of ontology (concepts and instances), fundamental for the construction of SWEETS. Figure 1 presents the SWEETS architecture.

The SWEETS architecture is divided into three layers. The first one (1) creates a lightweight ontology O_{ci}, which uses the co-occurrence of tags (terms) related to items (instances). The relationships between the terms of the lightweight ontology $O_{ci,}$ are weighted by the number of instances (I) which are tagged with its terms (tags). That is, the number of times that the terms co-occur in different instances. This is a basic method of text mining, in which terms are usually associated by its co-occurrence in documents (Mika, 2007; Feldman *et al.*, 1998; Cutting *et al.*, 1992).

When the lightweight ontology O_{ci} was generated, the second layer (2) is responsible to extract every concept that have more than 1 (one) relationship, their relationships and their weights. There is no agreement in which would be the best minimum for all cases. In this SWEETS version is used the minimum 4 (four) relationships. It is important to say that this number is configurable and therefore can be modified at any time.

After that, the information represented in the vector space $\overrightarrow{Wo_{t_i}} = [Wo_{t_1}, Wo_{t_2}, \ldots, Wo_{t_{n-1}}, Wo_{t_n}]$, where Wo_t represents the weight vector of the key concept and their relationships. The weight of the key concept is more relevant than

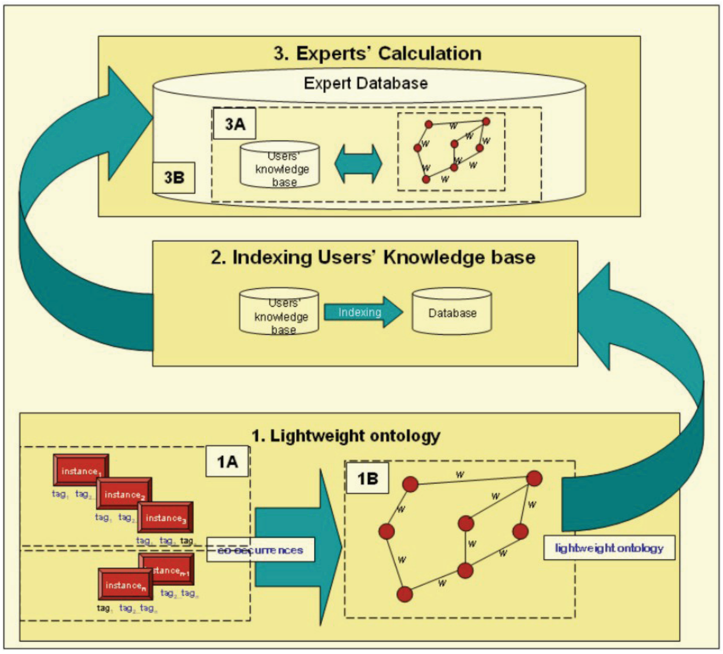

Figure 1. SWEETS architecture.

any other weight relationship, so this weight is determined by max($\overrightarrow{Wo}$)+1. So, for each concept and their relationships there will be a vector representation, composing a set of vectors. After this process the knowledge base of each user is indexed. This indexing is performed using all knowledge produced by users in the environment where SWEETS is being implemented. For each vector $\overrightarrow{Wo_t}$ there will be an equivalent vector $\overrightarrow{Fu_{t_i}}$ = [Fu_{t_1}, Fu_{t_2}, ... , $Fu_{t_{n-1}}$, Fu_{t_n}], where $\overrightarrow{Fu_t}$ represents the frequency of each user term and $\overrightarrow{Wo_t}$ represents the weights of terms in the lightweight ontology O_{ci}. It is important to highlight that the size of the vector $\overrightarrow{Wo_t}$ and the size of the vector $\overrightarrow{Fu_t}$ are the same, with Fu_{t_m}>=0 and Wo_{t_m}>=0.

Finally, the third layer calculates the users' expertise degree regarding the concepts. For this, we use the $\overrightarrow{Fu_t}$ and $\overrightarrow{Wo_t}$ representations. The calculation of the experts can be done by similarity analysis algorithms, like cosine, Pearson's coefficient and Jaccard (Baeza-Yates and Ribeiro, 1999). This version of SWEETS uses the cosine algorithm, as shown in Figure 2.

$$\mathrm{GEuc} = \frac{\sum_{i=1}^{n}\left(\mathrm{Wo}_{t_i} \cdot \mathrm{Fu}_{t_i}\right)}{\sqrt{\sum_{i=1}^{n}\left(\mathrm{Wo}_{t_i}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(\mathrm{Fu}_{t_i}\right)^2}}$$

Figure 2. Cosine Formula (Baeza-Yates and Ribeiro, 1999).

The GEuc function represents the user expertise degree in relation to a concept and its value varies between 0 and 1. The value 0 represents no expertise from the user regarding the concept, while the value 1 represents the total expertise of the user regarding the concept. According to Baeza-Yates and Ribeiro (1999), this function is inversely related to the angle between $\overrightarrow{Fu_t}$ and $\overrightarrow{Wo_t}$, because a smaller angle between $\overrightarrow{Fu_t}$ and $\overrightarrow{Wo_t}$, originates a greater cosine and a greater correlation between $\overrightarrow{Fu_t}$ and $\overrightarrow{Wo_t}$. Another characteristic of the cosine measure is its independency regarding the size of both vectors.

Social network environment

This research was conducted inside a Brazilian Innovation Institute called C.E.S.A.R., and, more specifically, inside its social network environment called a.m.i.g.o.s. In the next sections it is presented a brief introduction of C.E.S.A.R. and its knowledge management context, and introduction to a.m.i.g.o.s. social network.

C.E.S.A.R.

C.E.S.A.R. is a world-class private institution that develops products and processes, provides services, and cradles innovative new companies in their early stages using Information and Communication Technologies (ICT). It also provides innovation projects in many different areas. It is a 600-people organization associated with a Computing centre and with R&D departments from the private sector.

This institution is part of the "Porto Digital". An information and communication cluster with focus on software development, emphasis in innovation, entrepreneurship and human capital capture.

Like "Porto Digital", one of the main commitments of C.E.S.A.R. is innovation and it is all about knowledge building and sharing. The institute also works in collaboration with different University research groups in those cases where the needed expertise cannot be found within the organization. One of the important partners is the Federal University of Pernambuco, one of the top ranking universities in Brazil. Together their work accomplishes C.E.S.A.R.'s mission that is to transfer information technology knowledge between the industry and the academia in a self-sustainable way.

Currently C.E.S.A.R. is comprised of five buildings, four in Recife and one in São Paulo. It has more than 50 projects ranging from different areas like: Digital TV, Embedded Systems, Mobility, software reuse, etc. Each project size varies from 3 to 50 collaborators geographically distributed throughout its buildings and with regular communication.

Innovation; the array of different areas; buildings geographically distributed and the partnership with universities and the private sector are important aspects that make the need of a Knowledge Management tool vital for this institution. In environments like C.E.S.A.R., the KM tools must help collaborators to communicate more efficiently and help the institution to store the knowledge generated by them.

A.M.I.G.O.S.

In order to achieve a better result regarding its knowledge management strategy, C.E.S.A.R. has developed and incorporated the use of a WBSN tool named a.m.i.g.o.s. (Costa *et al.*, 2008). Its first version was deployed in October 2006. This new tool is helping C.E.S.A.R. to bring their employees and partners into a new virtual space for mutual cooperation improved communication, allowing the access and interchange of information and knowledge from anywhere. This new initiative is trying to add some human-oriented aspects to a system-oriented approach, reducing the gap to a dynamic-oriented approach.

Portuguese acronym for Multimedia Environment for Integration of Groups and Social Organizations, the main goal of a.m.i.g.o.s. is to provide a software infra-structure to support the creation of WBSN. The a.m.i.g.o.s. also intends to stimulate the knowledge creation and sharing by its members, and provides many features in order to be used as a knowledge sharing tool. Some of these features are presented on next sections.

Profiles

Each member has a personal profile. This profile is composed of a set of static information provided by user on registration, like physical address, languages, e-mail address, instant messengers' identifications and a brief personal description, focusing on interest areas.

However, the most important part of the user's profile is not filled explicitly by him/her, but inferred by the system. This information includes: (i) the user's activity index, which is calculated through the amount of activities that produce or consume knowledge on the environment; (ii) a set of subjects the user usually writes about, which are inferred through the identification of the most relevant terms posted on the environment by the user.

Stories and objects

Stories are intended to register, compile and disseminate emergent knowledge throughout all WBSN members. Any user can add their success or failure stories spontaneously. Each story can include different kinds of objects, as text files, slide presentations, audio or video files, so the knowledge, posted as text, can be enriched with media resources. In a similar fashion, a story can be related to other stories to allow users build bigger stories composed by many small stories.

Every member can also act as a reviewer of the content added by his peers, evaluating contributions qualitatively. This evaluation can be done in two ways: (i) by adding comments to stories, improving them with new knowledge and creating a dialogue around the added knowledge; (ii) giving a rate that ranges from 1 to 5 stars for stories. This range allows stories to be shown according to its relevance for the WBSN.

The system also allows adding knowledge through various kinds of objects. In the a.m.i.g.o.s. environment, every file that can store knowledge or be used to enrich existing knowledge is seen as an object, like text documents, papers, spreadsheets, audio and video files, and URLs to external resources. In order to increase user collaboration, any object added to the system can have comments, increasing the probability of surging new dialogues about the knowledge stored within objects.

Virtual communities

In the a.m.i.g.o.s. context, virtual communities can be seen as groups of users who have some interests in common. The system supports the creation of such communities by any WBSN member.

A community has three main mechanisms to support knowledge creation and sharing between its members. The main mechanism is the forum, where the members can start new topics about any interesting subject. A second mechanism is to associate stories to the community. A third one is to associate personal objects to the community. All mechanisms can be done by a community member only.

Folksonomy

To allow a classification of knowledge stored in a.m.i.g.o.s., facilitating the discovery of information by users, the a.m.i.g.o.s. have a folksonomy mechanism (Hansen *et al.*, 1999). With folksonomy, users can classify the content available in the environment of social and collaborative way, i.e., communities, histories, comments and discussion forums. For this only need to add keywords to the item being classified. Additionally, the system allows the visualization of all markers created by users through a tagcloud. Thus, the user can quickly access any content associated with a particular label (tag).

Preliminary experiment

Before implementing SWEETS and adding it to the a.m.i.g.o.s. environment, a questionnaire was applied to C.E.S.A.R.'s employees. The main objective of this questionnaire was to verify if there is a real need to add an expert recommender system in this environment.

The questionnaire was available at a.m.i.g.o.s. and was explicitly recommended to 100 users. Only 30 users answered the questionnaire, where 50% of them were software developers, 43.33% were systems analysts and 6.66% were project managers.

One of the main goals of the questionnaire was to investigate behaviors and attitudes of people when solving problems and how willing they are to help a colleague solving a problem. A question was made to understand the behavior of a user when solving a problem: "When you have a problem related to his work, which action would you take to solve the problem?" Figure 3 shows the results.

From all users, 26.67% answered that they look for other people they have direct access with to ask for help to solve a problem – that is, a colleague or a boss; 16.67% answered that they use books, manuals and other types of formal documentation; the remaining 56.67% try to solve the problem by searching on the web.

Users were also asked about how much they would be willing to help other people to solve a problem. With this question we found that 53.33% of people have always helped his colleagues on the last 5 times when help was requested, and only 10% of respondents have not helped – usually because they were really busy at the time or because they did not know the person who was asking for help.

These results indicate that it would be interesting for users to have an environment where people could cooperate to solve each other problems.

One incentive to foster this kind of collaboration could be by identifying users' expertise, as this information might help to find the right person to collaborate for the solution of the problem. These were some of the reasons

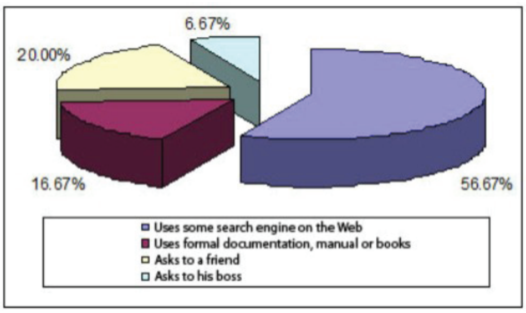

Figure 3. Behavior of users when solving a problem.

leading to the development of an expert recommender system and adding it to a.m.i.g.o.s.

The identification of experts in a.m.i.g.o.s. would be directly related to the production of knowledge by users, so this questionnaire included a question about that. As a result we found that only 20% of the users are used to produce and publish knowledge frequently, 33.3% of them are used to publish knowledge between 1 and 3 times per week, and most part of users – representing 46.66% of respondents – have rarely produced knowledge in a.m.i.g.o.s., because they believe that there are better ways to solve their problem other than using the a.m.i.g.o.s. social network environment.

As a way to solve this problem, or provide an incentive for people to produce and publish knowledge in a.m.i.g.o.s. was raised by the following question: "If there was an expert recommender system in a.m.i.g.o.s., would you feel more motivated to publish knowledge with a higher frequency?". 53.33% of users have responded "yes", while 36.67% said that they probably would, because they would like to use the tool first and then give their opinion and only 10% of people answered "no".

The majority of users who said they would feel more motivated to produce knowledge in a.m.i.g.o.s. argue that with an expert recommender system available, an improvement on the knowledge capture and publishing would be natural, as the employees would be interested in ascending in the enterprise, and this kind of system would be an excellent opportunity for employees to show their value and expertise. To C.E.S.A.R., the possibility of promoting more collaboration between people is important for both employees and enterprise, because the skills of employees would be better exploited, thereby increasing efficiency and effectiveness in the production level.

Experiment and results' implementation of the SWEETS at a.m.i.g.o.s.

The Expert Recommender System SWEETS, was developed and integrated into the a.m.i.g.o.s., a platform focused on knowledge management used in C.E.S.A.R. This integration with the a.m.i.g.o.s. had the following objectives:

(i) Use of the characteristics of social networks in a system that identifies domain experts;

(ii) Identify domain's experts;
(iii) Provide improvements in collaboration between users;
(iv) Improve the knowledge dissemination by users at a.m.i.g.o.s.;
(v) Improving the process of solving a problem;
(vi)And allocate or relocate people in projects with greater efficiency.

The majority of these goals, except for goals (i) and (ii), can be achieved with the analysis of results generated by SWEETS through time, e.g., 4, 5 or 6 months. With this information, comparative studies can be made using all information produced.

The analysis of the quality of recommendations generated by SWEETS was realized in two different ways. The first analysis used a limited group of 18 employees of C.E.S.A.R. This group was limited due to problems when finding people available to execute this analysis. However, we believe that this number is enough to examine the satisfaction level with the generated recommendations. The second analysis was applied only in version 2.0 because this version had an explicit way to identify the users' expertise. This analysis aimed at verifying the level of knowledge or interest of the users in a domain which they had been identified as specialist.

Results of the SWEETS

In the current version of SWEETS, users can explicitly request an expert through a keyword. This feature solves the problem of unnecessarily generated recommendations, when they are made without the explicit indication by the user that this expert is needed.

Folksonomy is used to generate lightweight ontology in SWEETS 2.0, aiming to make more effective recommendations. The folksonomy in a.m.i.g.o.s. is created with associations of tags to the items (tagging) made by users. These items are objects (a document or a site), stories and comments on stories, communities, discussion topics and messages in these topics. The lightweight ontology O_{ci} is composed by relationships among concepts (tags) associated with a weight. Both (relationship and weight) are determined by the co-occurrence of concepts in the items used to create the folksonomy. As higher the weight among the concepts, more often they co-occur, and greater is the semantic relationship between them. The main advantage of this mechanism is avoiding pre-determination of the ontology, as it emerge from users interaction in a.m.i.g.o.s., making this the ontology not limited to one particular domain.

To create or update this ontology requires a high computational cost, especially when the folksonomy is extensive. This is the case of the folksonomy in a.m.i.g.o.s. which has around 450 tags. Therefore, this process is executed within a 15 days interval, and at night, when the system is less used. Figure 4 shows the lightweight ontology emerged from a.m.i.g.o.s..

The lightweight ontology O_{ci} emergent in a.m.i.g.o.s. has 901 relationships. Therefore,

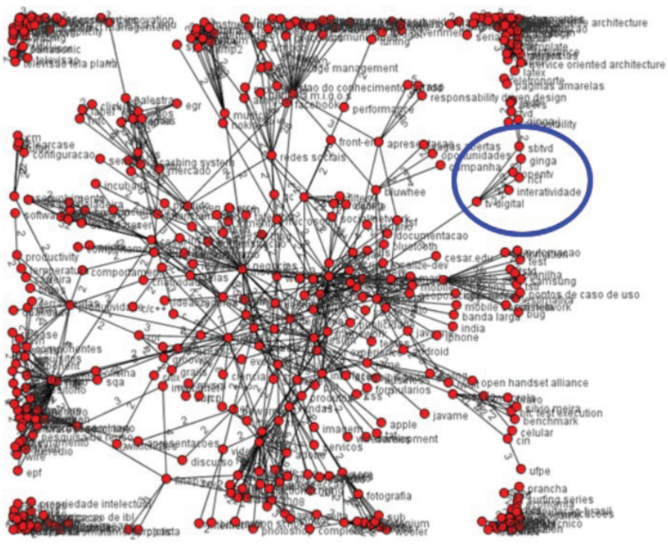

Figure 4. Lightweight ontology O_{ci} emerged in a.m.i.g.o.s.

the clutter of information shown in Figure 4 is not clear. In spite of that, an evidence of semantic in the ontology can be seen in the highlighted part. The ellipse emphasizes the relationships among terms “sbtvd”, “interactivity”, “ginga-j” and “opentv” with the term “digital tv”.

Some concepts that emerged from the folksonomy have similar meanings but are represented in different relationships on the lightweight ontology (e.g. “social networks” and “social network”). This problem was already expected, since the concepts (tags) are freely attached to items by users. Moreover, there may be typing errors. The description free is a feature of folksonomy, but these problems undermine the enrichment of the lightweight ontology, therefore, making the identification of experts less effective.

One of the ways to minimize these problems is to work with the awareness of users regarding the enrichment of the lightweight ontology, turning them into users willing to properly enrich the folksonomy. This work can be done making users aware that the folksonomy enrichment would be useful for both employees and company. If the system can correctly identify their skills, their potential can be better exploited, so their levels of satisfaction and productivity in the enterprise would rise.

The next step to infer the users’ expertise was indexing the knowledge base. While indexing users’ knowledge base, the concepts/terms of the lightweight ontology O_{ci} and their relationships are considered. It is important to note that not all concepts of the lightweight ontology are taken into account for every user, as the usage of these concepts is limited by the amount of relationships they have. The minimum number of relationships must be greater than 1. For this experiment, only concepts that have at least 4 relationships were counted. The reason for this choice is that, when the experiment was executed with a minimum of 3 relationships, a number of 387 experts were identified, from a total of 916 users – a number that probably would not be realistic, as approximately 50% of all users were considered as assiduous users of a.m.i.g.o.s. In a second moment, the experiment was executed considering the concepts that have at least 5 relationships. As a result only 60 experts were found. Several specialists in different areas were not represented on the final result. Therefore, the intermediate value, 4, is being used.

The expertise degree varies between 0 and 1. The level of expertise of a person should be greater than a certain threshold – configured on SWEETS – in order for him to be considered as an expert. The threshold used for this experiment was 0.8. Figure 5 shows an example of searching for a “requirements” expert at the user interface of the SWEETS 2.0 in the a.m.i.g.o.s. and the list of experts found.

By default, the list of recommended experts is sorted from highest to lowest expertise degree. Although not yet available in the interface, experts may be ordered by their availability (online or busy) or by its’ social distance to the user that is requesting the recommendation. For each expert recommended, it is possible to view the size of the path between the user and the recommended expert. It is also

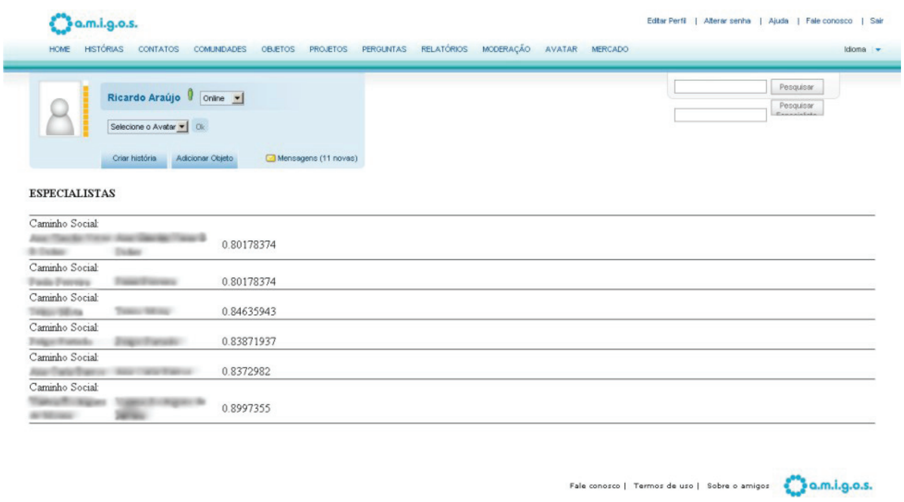

Figure 5. Interface of SWEETS 2.0 in a.m.i.g.o.s.

possible to see and navigate the full path of the social network from one person to another. Each name that composes the social path is a link to the profile of the respective user.

Analysis of the quality recommendations of SWEETS 2.0

The SWEETS 2.0 identified 121 experts in various domains. With the identification of experts, a customized questionnaire was developed and applied to approximately 50% of users in order to evaluate the knowledge level and the interest level in the areas in which each user was identified as expert.

The information gained with the application of this questionnaire was essential to evaluate the quality of recommendations generated by SWEETS 2.0. Avoiding a biased assessment, the user was not informed that he had been identified as an expert by the SWEETS 2.0.

The metric of accuracy was used to verify the quality of the recommendations. This metric is calculated by the ratio of relevant items and the total quantity of items. That is, the number of correctly identified experts related to the total detected experts. This experiment reached accuracy around 57%.The percentage of wrong experts was about 23%, while the remaining 20%, are probably justifiable, since it refers to subjects in which users have interest, however, they still aren't experts. This means that the users tend to produce knowledge on their interest subjects, one of the premises to detect expertise. Figure 6 shows the results of the questionnaire.

In SWEETS current version, there was good level of satisfaction shared by the majority of users regarding the possibility of searching for an expert using a specialty as the search parameter, avoiding unnecessary recommendations. During this research there was one employee who searched for a specialist in "Computer Networks" and got zero results. This happened because this area was not included in the lightweight ontology O_{ci}, so there is no way to identify such experts.

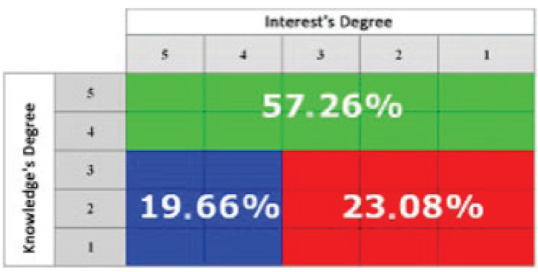

Figure 6. Recommendations' quality.

Conclusion

This paper presented an expert recommender system called SWEETS and its usage in a social network environment called a.m.i.g.o.s., a Web-Based Social Network (WBSN) focused on Knowledge Management, used as the main tool for knowledge creation, communication and dissemination at C.E.S.A.R., a Innovation Institute located in Brazil. Before the development of SWEETS, a survey was conducted among C.E.S.A.R's employees. That survey aimed to verify the real need for implementing an expert recommender system inside a.m.i.g.o.s. environment. According to 53.33% of the surveyed users, a tool like SWEETS would be so important to the environment that it would motivate them to a better engagement regarding knowledge production, as only 20% of users often produce knowledge inside environment. Building knowledge inside this kind of environment is fundamental to identify people's expertise.

In order to add semantics to recommendations and improve its quality, the SWEETS 2.0 was developed and added to a.m.i.g.o.s. It allows explicit search for experts and uses lightweight ontology as the knowledge representation mechanism. Unlike ICARE (Petry, 2007) this lightweight ontology is not a default one, it emerges from the social network folksonomy, and thus the identification of experts is domain free, which is the main advantage of the presented approach.

The usage of folksonomy as the basis for a lightweight ontology has some limitations. Due to the lack of control of how users build the folksonomy - this feature sometimes compromises the recommendations quality, because the users' entries may contain typing errors, spelling errors, or may cause some inconsistencies, as sometimes an user can apply a tag to an object (item), and other users can use other tags, grammatically different but with the same meaning.

In a survey conducted with users of the SWEETS 2.0, we observed that there was an improvement in the users' satisfaction. According to users, this improvement comes from the fact that the recommendations are not generated until it is needed, as users only receive recommendations of experts when they request it.

In the SWEETS 2.0, as the experts are explicitly recommended, a second analysis can be conducted to evaluate the precision level of the experts identified. This precision was 57.26%, while the percentage error was 23.08%. The remaining 19.66% can be justified, however, the users are not experts in such area (issues), they expressed some interests about them, and so there is a probability that users have posted relevant information on the environment about these areas.

To minimize the issues previously presented, some future work aims to apply a stemming algorithm (which reduces words to its radical) before generating the lightweight ontology, effectively improving recommendations quality. With this improvement, we expect to decrease the amount of redundancies of concepts represented in the lightweight ontology and expect a increasing on the semantic weight among these concepts. Furthermore, we also intend to enrich the lightweight ontology using a thesaurus, as WordNet (Fellbaum, 1998). For each term of lightweight ontology, it could be verified its existence in the thesaurus, its respective synonyms, hierarchies and other relationships among terms. In this case, all this information could be added to the lightweight ontology. Furthermore, we also intend to use machine learning methods and techniques for tag recommendation to the users, while these users are tagging objects (items) of the folksonomy.

Acknowledgement

This work was supported by the National Institute of Science and Technology for Software Engineering (INES - http://www.ines.org.br), funded by CNPq and FACEPE, grants 573964/2008-4 and APQ-1037-1.03/08.

References

BAEZA-YATES, R.; RIBEIRO, B. 1999. *Modern Information Retrieval*. New York, ACM Press, 544 p.

BOGARTTI, S.P.; CROSS, R. 2003. A Relational View of Information Seeking and Learning in Social Networks. *Management Science,* **49**(4):432-445. http://dx.doi.org/10.1287/mnsc.49.4.432.14428

BOOMEN, M. 2008. From Community Metaphor to WEB 2.0. *In:* WEB BASED COMMUNITIES, Amsterdam, 2008. *Proceedings...* Amsterdam, IADIS Press, p. 192-196.

BOSE, R. 2004. Knowledge management metrics. *Industrial management & data systems,* **104**:457-468.

CASONATO, R.; HARRIS, K. 2002. The Knowledge Worker Investment Paradox. Gartner Research. Available at: http://www.gartner.com/DisplayDocument?doc_cd=108326. Accessed on: 15/05/2011

CHOI, B.; LEE, H. 2003. An empirical investigation of KM styles and their effect on corporate performance. *Information & Management*, **40**(5):403-417.

COSTA, R.A.C.; OLIVEIRA, R.Y.S.; SILVA, E.M.S.; MEIRA, S.R.L.M. 2008. A.M.I.G.O.S.: Uma plataforma para Gestão de Conhecimento através de Redes Sociais. *In:* SIMPÓSIO BRASILEIRO DE SISTEMAS COLABORATIVOS IEEE COMPUTER SOCIETY, 2008, Washington, DC, 2008. *Proceedings...* Washington, SBSC, p. 192-203. http://dx.doi.org/10.1109/SBSC.2008.26

CUTTING, D.R.; KERGER, D.R.; PEDERSEN, J.O.; TUKEY, J.W. 1992. Scatter/Gather: a Cluster-based Approach to Browsing Large Document Collections. *In:* ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 15, Denmark, 1992. *Proceedings...* New York, ACM SIGIR, p. 318-329.

http://dx.doi.org/10.1145/133160.133214

DOMINGOS, P.; RICHARDSON, M. 2001. Mining the network value of customers. *In:* INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 7, San Francisco, 2001. *Proceedings...* New York, ACM Press, p. 57-66. http://dx.doi.org/10.1145/502512.502525

DUHON, B. 1998. It's all in our heads. *Inform*, **12(**8):8-13.

ERICKSON, T.; KELLOGG, W.A. 2000. Social Translucence: An Approach to Designing Systems that Support Social Processes, *ACM transactions on computer-human interaction (TOCHI)*, **7**:59-83 (Part 1).

EHRLICH, K.; LIN, C.Y.; FISCHER, V. F. 2007. Searching for experts in the enterprise: combining text and social network analysis. *In:* INTERNATIONAL ACM CONFERENCE ON SUPPORTING GROUP WORK, Sanibel Island, 2007. Proceedings... New York, ACM Press, p. 117-126. http://dx.doi.org/10.1145/1316624.1316642

FELDMAN, R.; DAGAN, I.; HIRSH, H. 1998. Mining Text Using Keyword Distributions. *Journal of Intelligent Information Systems*, **10**(3):281-300.

http://dx.doi.org/10.1023/A:1008623632443

FELLBAUM, C. 1998. *WordNet - An Electronic Lexical Database*. Cambridge, MIT Press, 423 p.

GAL-OZ, N.; GRINSHPOUN, T.; GUDES, E.; MEISELS, A. 2008. Cross-Community Reputation: Policies and Alternatives. *In*: WEB BASED COMMUNITIES, Amsterdam, 2008. *Proceedings...* Amsterdam, IADIS Press, p. 197-201.

HANSEN, M.; NOHRIA, N.; TIERNEY, T. 1999. What is your strategy for managing knowledge? *Harvard Business Review*, **77**(2):106-116.

IACOBUCCI, D.; WASSERMAN, S.; FAUST, K. 1994. *Social Network Analysis: methods and applications*. Cambridge, Cambridge University Press, 825 p.

KAUTZ, R.; SELMAN, B.; SHAH, M. 1997. ReferralWeb: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, **40**(3):63-65.

http://dx.doi.org/10.1145/245108.245123

LIN, C.Y.; EHRLICH, K.; FISCHER, V.G.; DESFORGES, C. 2008. SmallBlue: People Mining for Expertise Search. *Multimedia IEEE*, **15**(1):78-84.

http://dx.doi.org/10.1109/MMUL.2008.17

LIPNACK, J.; STAMPS, J. 1988. *The Networking Book: People Connecting with People*. New York, Routledge and Kegan Paul, 192 p.

MIKA, P. 2007. Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, **5**(1):5-15.

http://dx.doi.org/10.1016/j.websem.2006.11.002

NARDI, B.A.; WHITTAKER, S.; SCHWARZ, H. 2000. It's not what you know, it's who you know: Work in the information age. *First Monday*, **5**(5). Available at: http://firstmonday.org/issues/issue5_5/nardi/index.html. Accessed on: 01/10/2010.

NONAKA, I.; TAKEUCHI, H. 1995. *The knowledge creating company*. New York, Oxford University Press, 284 p.

PETRY, H. 2007. *ICARE: um sistema de recomendação de especialistas sensível a contexto.* Recife, PE. Dissertação de Mestrado. Universidade Federal de Pernambuco, p. 102.

PLICKERT, G.; CÔTÉA, R.R.; WELLMAN, B. 2007. It's not who you know, it's how you know them: Who exchanges what with whom? *Social Networks*, 29(3):405-429.

http://dx.doi.org/10.1016/j.socnet.2007.01.007

POLANYI, M. 1967. The tacit dimension. *In:* L. PRUSAK, *Knowledge in Organizations*. London, Butterworths, p. 135-146.

REICHLING, T.; SCHUBERT, K.; WULF, V. 2005. Matching Human Actors Based on their Texts: Design and Evaluation of an Instance of the Expert Finding Framework. *In:* INTERNATIONAL ACM SIGGROUP CONFERENCE ON SUPPORTING GROUP WORK, 2005. Proceedings, New York, ACM Press, p. 61-70.

http://dx.doi.org/10.1145/1099203.1099213

SILVA, E.M. 2009. *SWEETS: um sistema de recomendação de especialistas aplicado a redes sociais.* Recife, PE. Masters Dissertation. Universidade Federal de Pernambuco, 129 p.

STAAB, S.; DOMINGOS, P.; MIKA, P.; GOLBECK, J.; DING, L.; FININ, T.; JOSHI, A.; NOWAK, A.; VALLACHER, R.R. 2005. Social Networks Applied. *IEEE Intelligent Systems*, **20**:80-93.

http://dx.doi.org/10.1109/MIS.2005.16

YUKAWA, T.; KASAHARA, K.; KATO, T.; KITA, T. 2002. An Expert Recommendation System using Concept-based Relevance Discernment, Tools with Artificial Intelligence. *In:* INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 13, Dallas, 2002. *Proceedings...* Dallas, p. 257-264.

http://dx.doi.org/10.1109/ICTAI.2001.974472

Submitted on October 14, 2010.
Accepted on November 15, 2010.